

Ingest Dev Master Document

[Last updated](#)

[Owners](#)

[Why this document?](#)

[Other](#)

[Planning](#)

[Design](#)

[User Stories/UX](#)

[Ingest service](#)

[Architecture](#)

[API](#)

[Backup](#)

[Data](#)

[Database](#)

[Devops](#)

[Monitoring](#)

[Metadata](#)

[Reliability, Scalability and Test](#)

[Security](#)

[Submission](#)

[Pre-submission](#)

[Uploading](#)

[Envelope](#)

[Spreadsheet processing](#)

[Validating](#)

[Exporting](#)

[Updating](#)

[Accessioning and Archiving](#)

[UX](#)

[Metadata service](#)

[Datastore service](#)

[Terra](#)

[Analysis service](#)

[Query service](#)

[Data submission portal](#)

[EBI FIRE](#)

[EBI AAP](#)

[Globus](#)

[Wrangling](#)

[Releases](#)

Last updated

2020-06-05

Owners

Justin Clark-Casey

Why this document?

A living design master document that should have links to all other design documents.

Other

Documentation on development SOPs is in [the ingest-central wiki](#). SOPs that aren't there (e.g. are in Google docs linked here) should be moved there over time.

Wrangler SOPs live on [readthedocs](#).

Planning

[Ingest - Revised scoping Document](#)

[DCP - Strategy & Roadmapping](#)

[DCP Pilot Scope](#)

[HCA Plan for GA](#)

[Proposed ingest feature set roadmap](#)

[Ingest dev priorities](#)

[Ingest dev progress tracking](#)

[Ingest roadmap and product delivery strategy overview](#)

[Human Cell Atlas Data Coordination Platform - Our Strategy](#)

[The AIT HCA Roadmap - Strategy and Value - Feb 21st 2020](#)

[AIT HCA roadmap 2020 - DRAFT v2](#)

[HCA DCP 2.0 MVP Plan](#)

[Sprint planning agenda and rolling backlog](#)

[Embracing whole team Agile](#)

[Planning meeting Agenda 26 May 2020:](#)
[Ingest team pain points and ideas for new features](#)
[Ingest Sprint Milestones](#)
[Mid sprint check-in Agenda](#)
[Bi-weekly Ingest dev meeting](#)
[Technical Facilitator Charter](#)

Design

[DCP System architecture presentation](#)
[DCP Architecture and Data Flow Overview](#)
[DCP Terminology](#)
[Self led DCP tutorial](#)
[DCP Quarterly Goals Develop](#)
[DCP Quarterly Goals 1.0](#)
[Data model and life cycle pre-RFC](#)
[DCP Data Model Requirements](#)
[INSANE DCP reliability pre-RFC](#)
[Release notes](#)

User Stories/UX

[HCA DCP User Stories - Data Contribution - Master](#)
[HCA Platform - value proposition](#)
[The UX Design process - short intro and work on HCA](#) - as of 2019-10-11 Gabs says this is the most up to date doc showing user personas and stories
[HCA Personas and users](#)
[HCA Personas- Sharing Version](#)
[HCA User research summary](#)
[Beta 2 Feedback](#)
[HCA - MVP for data contribution](#)
[Contribution UX journey screens](#)

Ingest service

Architecture

[Ingest service charter](#)
[HCA Ingest service specification and overall design](#)
[Ingest Control Flows](#)
[Ingest Overview presentation](#)
[The Ingest Submission Domain Model](#)
[From submitter to archives - end to end](#)
[HCA DCP Terra Repo Ingest Meeting Agenda](#)

[DCP/2 : EBI Ingest to Terra Options Investigation](#)
[DCPv2 MVP Schema Design](#)
[DCP data flow periodicity](#)
[DCP2 \(meta\)data update design and scope](#)

API

[Ingest API example usage doc](#)

Backup

[Ingest Backup Testing Design](#)
[Ingest Backup Redaction Design](#)

Data

[Datasets to reingest prior to AUDR](#)
[CZI incentive important dates and deadlines](#)
[Making the Meyer dataset available](#)
[Dataset Tracking Sheet](#)

Database

[Ingest DB Clean-up Proposal 2019](#)
[Rebuilding the Ingest Database out of the DSS](#)

Devops

[SOPs](#)
[CI/CD in ingest](#)
[Ingest Gitlab CI/CD Integration Requirements](#)
[Ingest Service Deployment Via Gitlab](#)
[Cloud hosting setup for ingest services at the EBI](#)
[Setting up GitLab Runner](#)
[CI/CD Review](#)
[Upload Service alternatives](#)

Monitoring

[Ingest monitoring master document](#)

Metadata

[The Human Cell Atlas: Overview of Metadata Structure](#) (experimental graph)
[Metadata Submissions Tradeoffs](#)
[Ingest dev schema master document](#)

[Library prep entity/data groups meeting from 2019-09-09](#)
[Checksum Metadata Options](#)

Reliability, Scalability and Test

[Ingest Reliability and Scalability Master Document](#)
[DCP Scale Limits and Bottlenecks in Various Data Shapes](#)
[Ingest Test Master Document](#)
[Ingest traceability](#)
[Slow API call findings](#)

Security

[Ingest Auth* DCP/1 Status Document](#)
[Ingest Auth* HLTD](#)
[Ingest Auth* at EBI](#)
[Ingest Authz Technical Design](#)
[HCA Ingest to DSP auth management](#)

Submission

[Wrangling flow](#)
[Re-ingestion preserving project UUID](#)
[Submission flow](#)

Pre-submission

[Improve the submitter experience - stories and plans](#)
[Improve the submitter experience HLTD](#)

Uploading

[Submitting data files to the HCA Data Coordination Platform](#)
[EBI HCA Upload Service](#)
[The hca-util CLI tool](#)
[Genericising the hca-util CLI tool](#)

Envelope

[Submission Envelope Reporting](#)

Spreadsheet processing

[HCA spreadsheet spec 3.0](#)
[Array/module tab support in submission spreadsheets](#)

Validating

[Ingest file validation master document](#)

Exporting

[Bundling in ingest](#)

[Ingest Exporting Redesign 2019](#)

[Discussion: New Exporter Architecture](#)

[Data groups design](#)

[DCP2 exporter design](#)

Updating

[Ingest AUDR master document](#)

Accessioning and Archiving

The two are strongly linked.

[Accessioning service specification](#)

[Ingest archival specification](#)

[Pilot Ingest USI Status spreadsheet](#)

[USI Project Documentation](#)

[Archiving without the DataStore](#)

[Updating Archived Data Design](#)

[Ingestion to Archives Instructions](#)

[Upgrading the DSP to handle HCA Data](#)

[Mapping HCA Data to EBI Archive](#)

UX

[Ingest Status Indicator Proposal](#)

[Ingestion user process flow diagram](#)

Metadata service

[HCA Metadata Standard Overview](#)

[metadata-schema AIT instance](#)

Datastore service

[SOP: Retrieving bundles from the Data Store using project UUIDs](#)

Terra

[Using the Terra Data Repository](#)

Analysis service

[Handling simple metadata updates](#)

Query service

[Introduction to the Query Service](#)

Data submission portal

[Setting up a user account and logging in](#)

EBI FIRE

[FIRE infrastucture](#)

EBI AAP

[Authentication, Authorisation and Profile Service](#)

Globus

[GWAS_Deposition_Globus](#)

Wrangling

[EMBL-EBI Ingest Content and Wrangling Priorities](#)

[Creating and using upload areas](#) - the wranglers use these as holding areas for data before it is submitted

Releases

[HCA 20200331 release plan](#) - small 'r' release plan